# 第十二讲 约束最小二乘和$F$检验

2023.12.8

检验$q$个回归系数为0

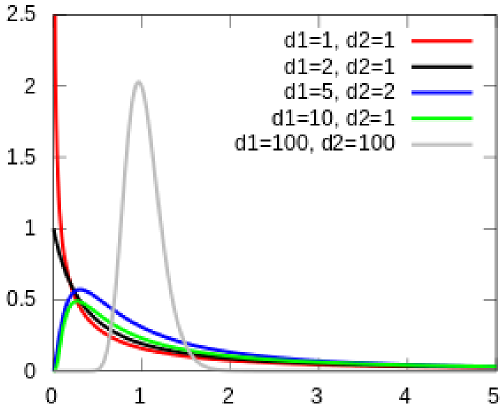
$$F=\frac{n-p}{q}\times\frac{R_p^2}{1-R_p^2}\overset{H_0}{\sim}F_{q,n-p}$$

# $F$分布

G.W. Snedecor (1881-1974), 美国统计学家，1947年在Iowa State Univ创建了美国第一个统计系。Snedecor为了表示对Fisher的敬意，称两个独立卡方随机变量的平均之比服从$F$分布，也叫做Snedecor's $F$-分布：

$$F=\frac{U/d_1}{V/d_2}\sim F_{d_1,d_2},\ \text{简记作}\ \frac{\chi^2_{d_1}/d_1}{\chi^2_{d_2}/d_2}=F_{d_1,d_2}$$

其中 $U\sim\chi^2_{d_1}, V\sim\chi^2_{d_2},\ U\perp\!\!\!\perp V.$




$$f(x)=\frac{\left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}}}{B\left(\frac{d_1}{2},\frac{d_2}{2}\right)}x^{\frac{d_1}{2}-1}\left(1+\frac{d_1}{d_2}x\right)^{-(d_1+d_1)/2}$$

注1: 若r.v. $x \sim t_d$，则 $x^2 \sim F_{1,d}$，简记作 $(t_d)^2 = F_{1,d}$或 $(t_d)^2 \sim F_{1,d}$。

注2： $F = \frac{U/d_1}{V/d_2} \sim F_{d_1,d_2}$，若$d_2 \to \infty$, 则由大数律，$V/d_2 \to 1$，因此

$$F \approx \frac{U}{d_1} \sim \chi^2_{d_1}/d_1,$$

若$d_1$也较大，则近似地$F \sim N(1,2/d_1)$ 。

注3: 若$F = \frac{U/d_1}{V/d_2} \sim F_{d_1,d_2}$（ $U \sim \chi^2_{d_1}, V \sim \chi^2_{d_2}$, $U \perp\!\!\!\perp V$ ），则

$$\frac{d_1 F}{d_1 F + d_2} = \frac{U}{U+V} \sim Beta(\frac{d_1}{2}, \frac{d_2}{2})$$

简记作 $\frac{\chi^2_{d_1}}{\chi^2_{d_1} + \chi^2_{d_2}} \sim Beta(\frac{d_1}{2}, \frac{d_2}{2})$

引理1．假设$\mathbf{x}\sim N_n(0,I_n)$，假设$B$是投影矩阵，则

（1）$\mathbf{x}^\top B\mathbf{x}=\|B\mathbf{x}\|^2\sim\chi_r^2, r=rank(B)$

（2）若$AB=0$, 则$A\mathbf{x}$与$B\mathbf{x}$独立，特别地，$A\mathbf{x}$与$\mathbf{x}^\top B\mathbf{x}$独立。

$B\mathbf{x}$的分布？

$B\mathbf{x}\sim N_n(0,B)$

退化多元正态分布

证明: 假设$B$的谱分解为$B=O\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}O^\top$, 令$\mathbf{y}=O^\top\mathbf{x}=\begin{pmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{pmatrix}\sim N(0,I_p)$,

(1) $\mathbf{x}^\top B\mathbf{x}=\mathbf{x}^\top O\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}O^\top\mathbf{x}=(\mathbf{y}_1{}^\top,\mathbf{y}_2{}^\top)\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{pmatrix}=\mathbf{y}_1{}^\top\mathbf{y}_1\sim\chi_r^2.$

(2) $B\mathbf{x}=O\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}O^\top\mathbf{x}=O\begin{pmatrix}\mathbf{y}_1\\ 0\end{pmatrix}=O_1\mathbf{y}_1$，其中$O_1$为$O=(O_1,O_2)$的前$r$列。

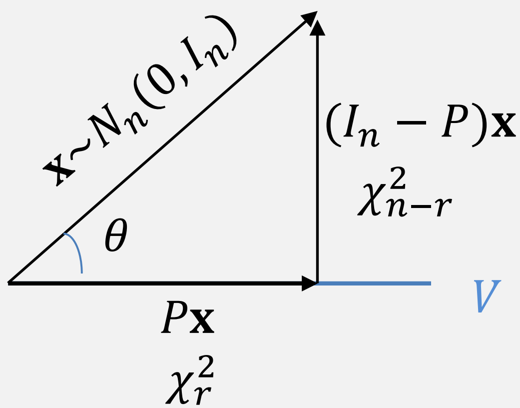
由$AB=0\Rightarrow AO\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}O^\top=0$，划分$AO=(C,D)\Rightarrow(C,D)\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}=0\Rightarrow C=0$,

则$A\mathbf{x}=AO\mathbf{y}=(C,D)\begin{pmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{pmatrix}=D\mathbf{y}_2$，与$B\mathbf{x}=O_1\mathbf{y}_1$独立。

命题1. 假设 $P=P_V$是一个(子空间$V$的)$n\times n$投影矩阵，$r=\mathrm{rank}(P)=\dim(V)$, 若 $\mathbf{x}\sim N_n(0,I_n)$，则

$$\frac{\|P\mathbf{x}\|^2/r}{\|(I_n-P)\mathbf{x}\|^2/(n-r)}\sim\frac{\chi_r^2/r}{\chi_{n-r}^2/(n-r)}\sim F_{r,n-r}$$

证： $\|P\mathbf{x}\|^2\sim\chi_r^2$，$\|(I_n-P)\mathbf{x}\|^2\sim\chi_{n-r}^2$, 且与 $\|P\mathbf{x}\|^2$独立。




| | | | | |
|---|---|---|---|---|
| $\mathbf{x}$ | $=$ | $P\mathbf{x}$ | $+$ | $(I_n-P)\mathbf{x}$ |
| $\|\mathbf{x}\|^2$ | $=$ | $\|P\mathbf{x}\|^2$ | $+$ | $\|(I_n-P)\mathbf{x}\|^2$ |
| $\chi_n^2$ | $=$ | $\chi_r^2$ | $+$ | $\chi_{n-r}^2$ |

$$\mathrm{ctg}^2(\theta)=\frac{\|P\mathbf{x}\|^2}{\|(I_n-P)\mathbf{x}\|^2}\sim\frac{r}{n-r}F_{r,n-r}$$

$$\cos^2(\theta)=\frac{\|P\mathbf{x}\|^2}{\|\mathbf{x}\|^2}\sim\mathrm{Beta}\left(\frac{r}{2},\frac{n-r}{2}\right)$$

当初定义F分布的时候，如果卡方不除以自由度，应用起来就会更方便一些。

# 一般线性假设的F检验

科学是一种系统性的知识体系，它积累并组织可检验的有关于万物的解释或理论。科学强调理论的可证伪性。科学不是寻求绝对的真理，而是在现有基础上，摸索式地不断接近真理。假设检验是对科学假设进行判断的工具。

Wald检验是最直观的检验。非正态总体假设下，其原假设下的渐近分布为卡方分布（我们直接使用该结论）。

一般的 Wald 检验

假设$\boldsymbol{\theta}$是参数向量，$\hat{\boldsymbol{\theta}}$是$\boldsymbol{\theta}$的一个估计(例如MLE、LS)，考虑零假设 $H_0:\varphi(\boldsymbol{\theta})=\mathbf{c}_0$ ($\mathbf{c}_0$是已知的$q\times 1$向量)，定义Wald统计量

$$W=(\varphi(\hat{\boldsymbol{\theta}})-\mathbf{c}_0)^{\mathrm{T}}\hat{\Sigma}^{-1}(\varphi(\hat{\boldsymbol{\theta}})-\mathbf{c}_0),$$

其中$\Sigma=\mathrm{var}(\varphi(\hat{\boldsymbol{\theta}}))$，$\hat{\Sigma}=\mathrm{var}(\varphi(\hat{\boldsymbol{\theta}}))$为$\Sigma$的估计。

$H_0$成立时，$W\xrightarrow{d}\chi_q^2,\ n\to\infty$。

当$\theta$是标量时，$H_0:\theta=\theta_0$，Wald检验统计量

$$Z=(\hat{\theta}-\theta_0)/sd(\hat{\theta}),\ \text{或}W=Z^2,$$

$H_0$成立时，$W\xrightarrow{d}\chi_1^2$或$Z\xrightarrow{d}N(0,1)$。

一般线性假设

假设正态线性回归模型：$\mathbf{y}_{n\times 1}=X_{n\times p}\boldsymbol{\beta}_{p\times 1}+\boldsymbol{\varepsilon}_{n\times 1}$，$\boldsymbol{\varepsilon}\sim N_n(0,\sigma^2 I_n)$

$H_0: A\boldsymbol{\beta}=\mathbf{c}_0$ 称为一般线性假设，其中$A,\mathbf{c}_0$已知，$A$行满秩。

有关$\boldsymbol{\beta}=(\beta_0,\beta_1,\ldots,\beta_{p-1})^{\mathrm{T}}$的检验问题一般都可写成上述线性假设的形式，如$p=3$时，$\boldsymbol{\beta}=(\beta_0,\beta_1,\beta_2)^{\mathrm{T}}$

$$H_0:\beta_1=0\Leftrightarrow H_0:\mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}=0,\mathbf{a}=(0,1,0)^{\mathrm{T}}$$

$$H_0:\beta_1=\beta_2=0\Leftrightarrow H_0:A\boldsymbol{\beta}=0,A=\begin{pmatrix}0&1&0\\0&0&1\end{pmatrix}$$

$A$的选取方式不唯一，例如上述第二个例子中也可取$A=\begin{pmatrix}0&1&0\\0&1&-1\end{pmatrix}$。

但后续一般线性假设的F检验是唯一的（不依赖于$A$的选取方式）。

**Wald 检验**

对于一般线性假设$H_0$： $A_{q\times p}\boldsymbol{\beta}=\mathbf{c}_0$， $\boldsymbol{\theta}=A\boldsymbol{\beta}$的LS估计为$\hat{\boldsymbol{\theta}}=A\hat{\boldsymbol{\beta}}$,

$\Sigma=\mathrm{var}(\hat{\boldsymbol{\theta}}\mid X)=\sigma^2A(X^\top X)^{-1}A^\top$,其估计$\hat{\Sigma}=\hat{\sigma}^2A(X^\top X)^{-1}A^\top$。

$H_0$的Wald检验(不假设误差正态):

$$W=(A\hat{\boldsymbol{\beta}}-\mathbf{c}_0)^\top\left(\hat{\sigma}^2A(X^\top X)^{-1}A^\top\right)^{-1}(A\hat{\boldsymbol{\beta}}-\mathbf{c}_0)\overset{H_0\text{下近似}}{\sim}\chi_q^2.$$

但线性模型的统计推断通常假设误差正态，此时$W$或$F=W/q$具有精确分布（F分布）。后续我们将假设误差正态。

**$F$检验**

当误差服从正态分布时，$\boldsymbol{\varepsilon}\sim N(0,\sigma^2I_n)$，定义$F$检验统计量

$$F=W/q,$$

我们将证明$F\overset{H_0}{\sim}F_{q,n-p}$， 或等价地$W\overset{H_0}{\sim}qF_{q,n-p}$.

正态线性模型

定理1. 假设模型 $\mathbf{y}_{n\times1} = X_{n\times p}\boldsymbol{\beta}_{p\times1} + \boldsymbol{\varepsilon}_{n\times1}$， $\boldsymbol{\varepsilon} \sim N(0,\sigma^2 I_n)$, $X$列满秩，则

(1) $\hat{\boldsymbol{\beta}} \mid X \sim N\left(\boldsymbol{\beta}, \sigma^2 (X^\top X)^{-1}\right)$,

(2) $\dfrac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}$，且$\hat{\boldsymbol{\beta}}, \hat{\sigma}^2$独立。

证明:

(1) 由$\mathbf{y} \mid X \sim N(X\boldsymbol{\beta}, \sigma^2 I_n)$知$\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbf{y} \sim N(\boldsymbol{\beta}, \sigma^2 (X^\top X)^{-1})$。

(2) 因为$\mathbf{e} = (I_n - P_X)\mathbf{y} = (I_n - P_X)(X\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = (I_n - P_X)\boldsymbol{\varepsilon}$,

所以 $(n-p)\hat{\sigma}^2/\sigma^2 = \|\mathbf{e}\|^2/\sigma^2 = \mathbf{z}^\top (I_n - P_X)\mathbf{z}$, 其中$\mathbf{z} = \boldsymbol{\varepsilon}/\sigma \sim N(0, I_n)$,

$rank(I_n - P_X) = tr(I_n - P_X) = n - p$， 所以$(n-p)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-p}$.

另外，$\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbf{y} = (X^\top X)^{-1} X^\top (X\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (X^\top X)^{-1} X^\top \boldsymbol{\varepsilon} \triangleq \boldsymbol{\beta} + B\boldsymbol{\varepsilon}$

$\mathbf{e} = (I_n - P_X)\boldsymbol{\varepsilon}$, 由$B\ (I_n - P_X) = 0$,知$\hat{\boldsymbol{\beta}}$与$\hat{\sigma}^2$独立。

$\mathbf{x} \sim N(0, \sigma^2 I_n)$,若$AB^\top = 0$,则$A\mathbf{x}, B\mathbf{x}$独立.

定理2. 线性模型$\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon},\boldsymbol{\varepsilon}\sim N(0,\sigma^2 I_n)$, 假设$A_{q\times p}$是已知的行满秩矩阵

则 $F=\dfrac{(A\hat{\boldsymbol{\beta}}-\mathbf{c}_0)^\top\left(A(X^\top X)^{-1}A^\top\right)^{-1}(A\hat{\boldsymbol{\beta}}-\mathbf{c}_0)}{q\hat{\sigma}^2}\overset{\mathrm{H}_0}{\sim}F_{q,n\text{-}p}$。

证明：$\hat{\boldsymbol{\beta}}\,|\,X\sim N(\boldsymbol{\beta},\sigma^2(X^\top X)^{-1})\Rightarrow A\hat{\boldsymbol{\beta}}\sim N\left(A\boldsymbol{\beta}-\mathbf{c}_0,\sigma^2 A(X^\top X)^{-1}A^\top\right)$,

故$H_0$成立时，$\mathbf{z}\overset{\Delta}{=}\left(A(X^\top X)^{-1}A^\top\right)^{-1/2}(A\hat{\boldsymbol{\beta}}-A\boldsymbol{\beta})/\sigma\sim N(0,I_q)$，从而

$$a\overset{\Delta}{=}\|\mathbf{z}\|^2=\frac{1}{\sigma^2}(A\hat{\boldsymbol{\beta}}-A\boldsymbol{\beta})^\top\left(A(X^\top X)^{-1}A^\top\right)^{-1}(A\hat{\boldsymbol{\beta}}-A\boldsymbol{\beta})\sim\chi^2_q,$$

由定理1, $b\overset{\Delta}{=}\dfrac{1}{\sigma^2}(n-p)\hat{\sigma}^2\sim\chi^2_{n-p}$, 且$\hat{\sigma}^2$与$\hat{\boldsymbol{\beta}}$独立, $\Rightarrow a,b$独立，故$H_0$成立时

$$\frac{a/q}{b/(n-p)}=\frac{(A\hat{\boldsymbol{\beta}}-A\boldsymbol{\beta})^\top\left(A(X^\top X)^{-1}A^\top\right)^{-1}(A\hat{\boldsymbol{\beta}}-A\boldsymbol{\beta})}{q\hat{\sigma}^2}\sim F_{q,n-p}$$

推论：考虑一般线性假设$H_0: A\boldsymbol{\beta}=\mathbf{c}_0$, 其中$A_{q\times p},\mathbf{c}_{0\,q\times 1}$已知，则$H_0$成立时,

$$F=(A\hat{\boldsymbol{\beta}}-\mathbf{c}_0)^\top\left(A(X^\top X)^{-1}A^\top\right)^{-1}(A\hat{\boldsymbol{\beta}}-\mathbf{c}_0)/q\hat{\sigma}^2\sim F_{q,n\text{-}p}。$$

# 约束最小二乘

下面从投影角度考察一般线性假设的$F$检验。

首先考虑零假设$H_0: A\boldsymbol{\beta}=\mathbf{c}_0$成立时$\boldsymbol{\beta}$的约束最小二乘估计。

我们不妨假设$\mathbf{c}_0=0$，这是因为：

$A\boldsymbol{\beta}=\mathbf{c}_0 \Leftrightarrow A\boldsymbol{\beta}^*=0$，其中$\boldsymbol{\beta}^*=\boldsymbol{\beta}-A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}\mathbf{c}_0$，此时模型

$$\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon}=X\left(\boldsymbol{\beta}^*+A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}\mathbf{c}_0\right)+\boldsymbol{\varepsilon}=X\boldsymbol{\beta}^*+XA^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}\mathbf{c}_0+\boldsymbol{\varepsilon}$$

记$\mathbf{y}^*=\mathbf{y}-XA^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}\mathbf{c}_0$（注意$XA^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}\mathbf{c}_0$已知），则

$$\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon},\ H_0:\ A\boldsymbol{\beta}=\mathbf{c}_0 \Leftrightarrow \mathbf{y}^*=X\boldsymbol{\beta}^*+\boldsymbol{\varepsilon},\ H_0:\ A\boldsymbol{\beta}^*=\mathbf{0}$$

约束最小二乘问题

假设线性模型$\mathbf{y}=X_{n\times p}\boldsymbol{\beta}+\boldsymbol{\varepsilon},\quad \boldsymbol{\varepsilon}\sim(0,\sigma^2 I_n)$，设$A_{q\times p}$为行满秩常数矩阵，线性约束下的最小二乘问题

$$\min_{\boldsymbol{\beta}\in R^p}\|\mathbf{y}-X\boldsymbol{\beta}\|^2\quad s.t.\quad A\boldsymbol{\beta}=\mathbf{0}\qquad (*)$$

命题2.假设$X$列满秩，约束最小二乘问题（*）的最优解，即约束最小二乘估计为

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - (X^\top X)^{-1} A^\top \left(A(X^\top X)^{-1} A^\top\right)^{-1} A\hat{\boldsymbol{\beta}}.$$

证明1: 应用拉格朗日乘子法, 令

$$f(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \| \mathbf{y} - X\boldsymbol{\beta} \|^2 + 2\boldsymbol{\lambda}^\top (A\boldsymbol{\beta} - \mathbf{c}_0)$$

对$\boldsymbol{\beta}$求偏导并令之为0得：$X^\top X\boldsymbol{\beta} - X^\top \mathbf{y} + A^\top \boldsymbol{\lambda} = 0$

$\Rightarrow \tilde{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbf{y} - (X^\top X)^{-1} A^\top \boldsymbol{\lambda} = \hat{\boldsymbol{\beta}} - (X^\top X)^{-1} A^\top \boldsymbol{\lambda}$

代入$0 = A\tilde{\boldsymbol{\beta}} = A\hat{\boldsymbol{\beta}} - A(X^\top X)^{-1} A^\top \boldsymbol{\lambda} \Rightarrow \boldsymbol{\lambda} = \left(A(X^\top X)^{-1} A^\top\right)^{-1} A\hat{\boldsymbol{\beta}}$,

$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - (X^\top X)^{-1} A^\top \left(A(X^\top X)^{-1} A^\top\right)^{-1} A\hat{\boldsymbol{\beta}}.$

注： 由命题1，在线性约束下，拟合值

$\hat{\mathbf{y}}_0 = X\tilde{\boldsymbol{\beta}} = \hat{\mathbf{y}} - X(X^\top X)^{-1} A^\top \left(A(X^\top X)^{-1} A^\top\right)^{-1} A\hat{\boldsymbol{\beta}}$

这应该是$\mathbf{y}$在约束下的自变量空间$V_0$的投影，$V_0 = ?$

命题3. 约束$A_{q\times p}\boldsymbol{\beta}=\mathbf{0}$成立时，$X$张成的空间（称为自变量空间）

$$V_0 \overset{\Delta}{=} \left\{X\boldsymbol{\beta}: A\boldsymbol{\beta}=0,\ \boldsymbol{\beta}\in R^p\right\} = C(X(X^\top X)^{-1}A^\top)^\perp \cap C(X)$$

且该空间的投影阵$P_{V_0}=P_X-P_{X(X^\top X)^{-1}A^\top}$。

证明：若$\boldsymbol{\mu}\in V_0=\left\{X\boldsymbol{\beta}: A\boldsymbol{\beta}=0, \boldsymbol{\beta}\in R^p\right\}\subset C(X)$,

即存在$\boldsymbol{\beta}\in R^p$使得$\boldsymbol{\mu}=X\boldsymbol{\beta}\Rightarrow\ \boldsymbol{\beta}=(X^\top X)^{-1}X^\top\boldsymbol{\mu}$,

$\mathbf{0}=A\boldsymbol{\beta}=A(X^\top X)^{-1}X^\top\boldsymbol{\mu}\ \Rightarrow\ \boldsymbol{\mu}\in C(X(X^\top X)^{-1}A^\top)^\perp$，但$\boldsymbol{\mu}\in C(X)$,

所以$\boldsymbol{\mu}\in C(X)\cap C(X(X^\top X)^{-1}A^\top)^\perp$，所以

$$V_0\subset C(X)\cap C(X(X^\top X)^{-1}A^\top)^\perp。$$

反之也对，所以$V_0=C(X)\cap C(X(X^\top X)^{-1}A^\top)^\perp$,

即$C(X(X^\top X)^{-1}A^\top)\oplus V_0=C(X)$,所以$P_X=P_{X(X^\top X)^{-1}A^\top}+P_{V_0}$.

注：$\mathbf{y}$在$V_0$上的投影$\hat{\mathbf{y}}_0=P_{V_0}\mathbf{y}$使得约束最小二乘(*)达到最小，

$$\hat{\mathbf{y}}_0=P_{V_0}\mathbf{y}=P_V\mathbf{y}-P_{X(X^\top X)^{-1}A^\top}\mathbf{y}=X\hat{\boldsymbol{\beta}}-X(X^\top X)^{-1}A^\top\left(A(X^\top X)^{-1}A^\top\right)^{-1}A\hat{\boldsymbol{\beta}}$$

$$=X[\hat{\boldsymbol{\beta}}-(X^\top X)^{-1}A^\top\left(A(X^\top X)^{-1}A^\top\right)^{-1}A\hat{\boldsymbol{\beta}}]=X[\tilde{\boldsymbol{\beta}}]$$

由此我们用投影方法求出了$\boldsymbol{\beta}$的约束LS估计$\tilde{\boldsymbol{\beta}}$.

需要说明的是，理论上我们需要命题2的约束最小二乘估计的数学公式，但在具体的线性假设检验问题中可以将约束代入模型进行简化。

例1. 假设模型

$$\mathbf{y} = \mathbb{1}\beta_0 + \mathbf{x}_1\beta_1 + \cdots + \mathbf{x}_{p-1}\beta_{p-1} + \boldsymbol{\varepsilon}$$

设计阵$X = C(\mathbb{1}, \mathbf{x}_1, \ldots, \mathbf{x}_{p-1})$是$n \times p$矩阵, 约束: $\beta_1 = \cdots = \beta_{p-1}$

将约束代入模型得

$$\mathbf{y} = \mathbb{1}\beta_0 + (\mathbf{x}_1 + \cdots \mathbf{x}_{p-1})\beta_1 + \boldsymbol{\varepsilon},$$

设计阵$X_0 = C(\mathbb{1},\ \mathbf{x}_1 + \ldots + \mathbf{x}_{p-1})$是$n \times 2$ 矩阵, 这是简单线性模型，由此解得$\beta_1$的LS估计$\hat{\beta}_1$,

$\Rightarrow \beta_1, \ldots, \beta_{p-1}$的约束 LS估计都是 $\hat{\beta}_1$。

# $F$检验的其它表示

记权模型下和零模型（null model, 零假设下的模型）下的投影分别为

$$\hat{\mathbf{y}} = P_V \mathbf{y} = X\hat{\boldsymbol{\beta}},\ \ \hat{\mathbf{y}}_0 = P_{V_0}\mathbf{y} = X\tilde{\boldsymbol{\beta}}$$

其中$V = \mathrm{C}(X)$， $V_0 = \{X\boldsymbol{\beta} | A\boldsymbol{\beta} = \mathbf{0}\}$, 则由命题2知

$$\hat{\mathbf{y}} - \hat{\mathbf{y}}_0 = X(X^\top X)^{-1} A^\top [A(X^\top X)^{-1} A^\top]^{-1} A\hat{\boldsymbol{\beta}},$$

$$\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2 = \hat{\boldsymbol{\beta}}^\top A^\top [A(X^\top X)^{-1} A^\top]^{-1} A\hat{\boldsymbol{\beta}}$$

代表全模型和零模型的拟合值差别，，是$F$检验的关键部分。由定理2，

$$F = \hat{\boldsymbol{\beta}}^\top A^\top [A(X^\top X)^{-1} A^\top]^{-1} A\hat{\boldsymbol{\beta}} / q\hat{\sigma}^2 = \frac{n-p}{q} \times \frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2}{\|\mathbf{y} - \hat{\mathbf{y}}\|^2}$$

进一步，根据几何和统计意义我们还可得到$F$统计量的其它表示。 例如

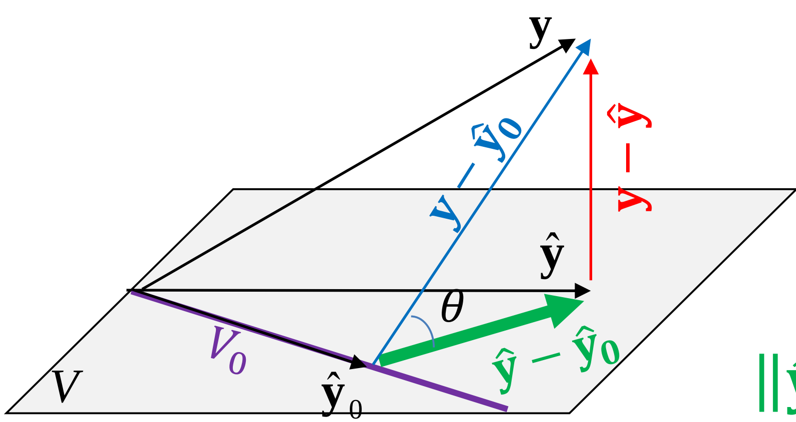
$$\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}_0\|^2 - \|\mathbf{y} - \hat{\mathbf{y}}\|^2 = RSS_0 - RSS$$

度量了全模型和零模型的残差平方和的差别

$$\Rightarrow\ F = \frac{n-p}{q} \times \frac{RSS_0 - RSS}{RSS}$$

# 对比全模型和零模型

| | 自变量空间 | 投影 | 残差平方和 | 决定系数 |
|---|---|---|---|---|
| 全模型 $\boldsymbol{\beta} \in R^p$ | $V = \{X\boldsymbol{\beta} \mid \boldsymbol{\beta} \in R^p\}$ $\dim(V) = p$ | $\hat{\mathbf{y}} = P_V\mathbf{y}$ | $RSS = \|\mathbf{y} - \hat{\mathbf{y}}\|^2$ | $R^2 = \dfrac{\|\hat{\mathbf{y}} - \mathbb{1}\bar{y}\|^2}{\|\mathbf{y} - \mathbb{1}\bar{y}\|^2}$ |
| 零模型 $A\boldsymbol{\beta} = \mathbf{0}$ | $V_0 = \{X\boldsymbol{\beta} \mid A\boldsymbol{\beta} = \mathbf{0}\}$ $\dim(V_0) = p - q$ | $\hat{\mathbf{y}}_0 = P_{V_0}\mathbf{y}$ | $RSS_0 = \|\mathbf{y} - \hat{\mathbf{y}}_0\|^2$ | $R_0^2 = \dfrac{\|\hat{\mathbf{y}}_0 - \mathbb{1}\bar{y}\|^2}{\|\mathbf{y} - \mathbb{1}\bar{y}\|^2}$ |




$$
\begin{aligned}
\mathbf{y} - \hat{\mathbf{y}}_0 &= \hat{\mathbf{y}} - \hat{\mathbf{y}}_0 + \mathbf{y} - \hat{\mathbf{y}} \\
\|\mathbf{y} - \hat{\mathbf{y}}_0\|^2 &= \|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2 + \|\mathbf{y} - \hat{\mathbf{y}}\|^2 \\
\chi^2_{n-\dim(V_0)} &= \chi^2_{\dim(V)-\dim(V_0)} + \chi^2_{n-\dim(V)}
\end{aligned}
$$

$\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2$：全模型和零模型的差别

$$
F = \frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2/q}{\|\mathbf{y} - \hat{\mathbf{y}}\|^2/(n-p)} = \frac{(RSS_0 - RSS)/q}{RSS/(n-p)}
$$

$$
R_p^2 = cos^2(\theta) = \frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2}{\|\mathbf{y} - \hat{\mathbf{y}}_0\|^2} = \frac{RSS_0 - RSS}{RSS_0} = \frac{R^2 - R_0^2}{1 - R_0^2}
$$

$$
F = \frac{n-p}{q} \times \frac{R^2 - R_0^2}{1 - R^2} = \frac{n-p}{q} \times \frac{R_p^2}{1 - R_p^2}
$$

细节参见命题4

命题4. 线性模型$\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon},\boldsymbol{\varepsilon}\sim N(0,\sigma^2 I_n)$，$H_0: A\boldsymbol{\beta}=\mathbf{0}$的$F$检验

(1) $$F=\frac{(A\hat{\boldsymbol{\beta}})^{\top}\left(A(X^{\top}X)^{-1}A^{\top}\right)^{-1}(A\hat{\boldsymbol{\beta}})}{q\hat{\sigma}^2}=\frac{n-p}{q}\times\frac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2}{\|\mathbf{y}-\hat{\mathbf{y}}\|^2}\sim F_{q,n-p}$$

(2) $F$有如下等价表示： $$F=\frac{n-p}{q}\times\frac{(RSS_0-RSS)}{RSS}$$

(3)若$A=(\mathbf{0},*)$,即$A$的第一列为0，则 $$F=\frac{n-p}{q}\times\frac{R^2-R_0^2}{1-R^2}=\frac{n-p}{q}\times\frac{R_p^2}{1-R_p^2},$$

其中$R_p^2=\dfrac{R^2-R_0^2}{1-R_0^2}=\dfrac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2}{\|\mathbf{y}-\hat{\mathbf{y}}_0\|^2}$。

前两页已基本证明，可忽略

证明：(1) 零假设成立时，$\boldsymbol{\beta}$满足约束$A\boldsymbol{\beta}=\mathbf{0}$,由命题3,

$$\hat{\mathbf{y}}_0=P_{V_0}\mathbf{y}=P_V\mathbf{y}-P_{X(X^{\top}X)^{-1}A^{\top}}\mathbf{y}=\hat{\mathbf{y}}-X(X^{\top}X)^{-1}A^{\top}\left(A(X^{\top}X)^{-1}A^{\top}\right)^{-1}A\hat{\boldsymbol{\beta}}$$

$\Rightarrow\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2=\hat{\boldsymbol{\beta}}^{\top}A^{\top}\left(A(X^{\top}X)^{-1}A^{\top}\right)^{-1}A\hat{\boldsymbol{\beta}}$，而$\hat{\sigma}^2=\|\mathbf{y}-\hat{\mathbf{y}}\|^2/(n-p)$,

$$\Rightarrow F=\frac{\hat{\boldsymbol{\beta}}^{\top}A^{\top}\left(A(X^{\top}X)^{-1}A^{\top}\right)^{-1}A\hat{\boldsymbol{\beta}}}{q\hat{\sigma}^2}=\frac{n-p}{q}\times\frac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2}{\|\mathbf{y}-\hat{\mathbf{y}}\|^2}。$$

下面利用投影表示证明$F \overset{H_0}{\sim} F_{q,n-p}$(比定理2的证明更直观一些):

$H_0$下$A\boldsymbol{\beta}=\mathbf{0}$, 故$P_{X(X^\top X)^{-1}A^\top} X\boldsymbol{\beta}=(\cdots)A\boldsymbol{\beta}=\mathbf{0}$,

所以 $\hat{\mathbf{y}}-\hat{\mathbf{y}}_0=(P_V-P_{V_0})\mathbf{y}=P_{X(X^\top X)^{-1}A^\top}(X\boldsymbol{\beta}+\boldsymbol{\varepsilon})=P_{X(X^\top X)^{-1}A^\top}\boldsymbol{\varepsilon}$,

由推论1, $a \overset{\Delta}{=} \|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2 \sim \sigma^2\chi_q^2$,其中$rank(X(X^\top X)^{-1}A^\top = rank(A)=q$.

另一方面$\mathbf{y}-\hat{\mathbf{y}}=(I-P_V)(X\boldsymbol{\beta}+\boldsymbol{\varepsilon})=(I-P_V)\boldsymbol{\varepsilon} \quad \Rightarrow b \overset{\Delta}{=} \|\mathbf{y}-\hat{\mathbf{y}}\|^2 \sim \sigma^2\chi_{n-p}^2$

因为$(I-P_V)P_{X(X^\top X)^{-1}A^\top}=0$,$a,b$独立,所以$H_0$下$F=\dfrac{a/q}{b/(n-p)} \sim F_{q,n-p}$.

(2) 因为$\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2=\|\mathbf{y}-\hat{\mathbf{y}}_0\|^2-\|\mathbf{y}-\hat{\mathbf{y}}\|^2=RSS_0-RSS \Rightarrow F=\dfrac{(RSS_0-RSS)/q}{RSS/(n-p)}$

(3) 记$A=(\mathbf{0},B)$,则$A\boldsymbol{\beta}=(\mathbf{0},B)\begin{pmatrix}\beta_0\\ \mathbf{b}\end{pmatrix}=B\mathbf{b}$(从而约束$A\boldsymbol{\beta}=0$对$\beta_0$没有约束),

而$X\boldsymbol{\beta}=(\mathbf{1},\mathrm{Z})\begin{pmatrix}\beta_0\\ \mathbf{b}\end{pmatrix}=\mathbf{1}\beta_0+Z\mathbf{b} \Rightarrow V_0=\left\{X\boldsymbol{\beta}:A\boldsymbol{\beta}=0,\boldsymbol{\beta}\in R^p\right\}=\left\{\mathbf{1}\beta_0+Z\mathbf{b}:B\mathbf{b}=0,\mathbf{b}\in R^{p-1}\right\}$,

$$\mathbf{1} \in V_0 = \left\{X\boldsymbol{\beta} : A\boldsymbol{\beta} = 0,\ \boldsymbol{\beta} \in R^p\right\} = \left\{\mathbf{1}\beta_0 + Z\mathbf{b} : B\mathbf{b} = 0, \mathbf{b} \in R^{p-1}\right\},$$

所以 $\mathbf{1} \perp \hat{\mathbf{y}} - \hat{\mathbf{y}}_0$，$\mathbf{1}^{\mathsf{T}}\hat{\mathbf{y}}_0 = \mathbf{1}^{\mathsf{T}}\hat{\mathbf{y}} = \mathbf{1}^{\mathsf{T}}\mathbf{y} = n\bar{y}$

$$R^2 = \frac{\|\hat{\mathbf{y}} - \mathbf{1}\bar{y}\|^2}{\|\mathbf{y} - \mathbf{1}\bar{y}\|^2}, R_0^2 = \frac{\|\hat{\mathbf{y}}_0 - \mathbf{1}\bar{y}\|^2}{\|\mathbf{y} - \mathbf{1}\bar{y}\|^2}，\|\hat{\mathbf{y}}_0 - \mathbf{1}\bar{y}\|^2 + \|\hat{\mathbf{y}}_0 - \mathbf{1}\bar{y}\|^2 = \|\hat{\mathbf{y}} - \mathbf{1}\bar{y}\|^2$$

$$R^2 - R_0^2 = \frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2}{\|\mathbf{y} - \mathbf{1}\bar{y}\|^2}$$

而$1 - R^2 = 1 - \frac{\|\hat{\mathbf{y}} - \mathbf{1}\bar{y}\|^2}{\|\mathbf{y} - \mathbf{1}\bar{y}\|^2} = \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|^2}{\|\mathbf{y} - \mathbf{1}\bar{y}\|^2}$，所以

$$\frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2}{\|\mathbf{y} - \hat{\mathbf{y}}\|^2} = \frac{R^2 - R_0^2}{1 - R^2} \Rightarrow F = \frac{n-p}{q} \times \frac{R^2 - R_0^2}{1 - R^2}。$$

若记$R_p^2 = \frac{R^2 - R_0^2}{1 - R_0^2}$，则$F = \frac{n-p}{q} \times \frac{R_p^2}{1 - R_p^2}$. 证毕

# 部分回归系数的$F$检验

模型：$\mathbf{y}_{n\times1}=X_{n\times p}\boldsymbol{\beta}_{p\times1}+\boldsymbol{\varepsilon}=\mathbf{1}\beta_0+X_1\boldsymbol{\beta}_1+X_2\boldsymbol{\beta}_2+\boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon}\sim N(0,\sigma^2 I_n),\ V=C(\mathbf{1},X_1,X_2)$

原假设 $H_0:\boldsymbol{\beta}_2=\mathbf{0}_{q\times1}$，记$H_0$成立时的子模型为$\mathbf{y}=\mathbf{1}\beta_0+X_1\boldsymbol{\beta}_1+\boldsymbol{\varepsilon},\ V_0=C(\mathbf{1},X_1)$。

命题5. 假设模型 $\mathbf{y}_{n\times1}=X_{n\times p}\boldsymbol{\beta}_{p\times1}+\boldsymbol{\varepsilon}=\mathbf{1}\beta_0+X_1\boldsymbol{\beta}_1+X_2\boldsymbol{\beta}_2+\boldsymbol{\varepsilon}$，$H_0$: $\boldsymbol{\beta}_2=\mathbf{0}_{q\times1}$，

记$X_2^{\perp}=X_2-P_{V_0}X_2=X_2-P_{\mathbf{1},X_1}X_2$，则$H_0$的$F$检验统计量为

$$F=\frac{\|X_2^{\perp}\hat{\boldsymbol{\beta}}_2\|^2}{q\hat{\sigma}^2}\overset{H_0}{\sim}F_{q,n-p},$$

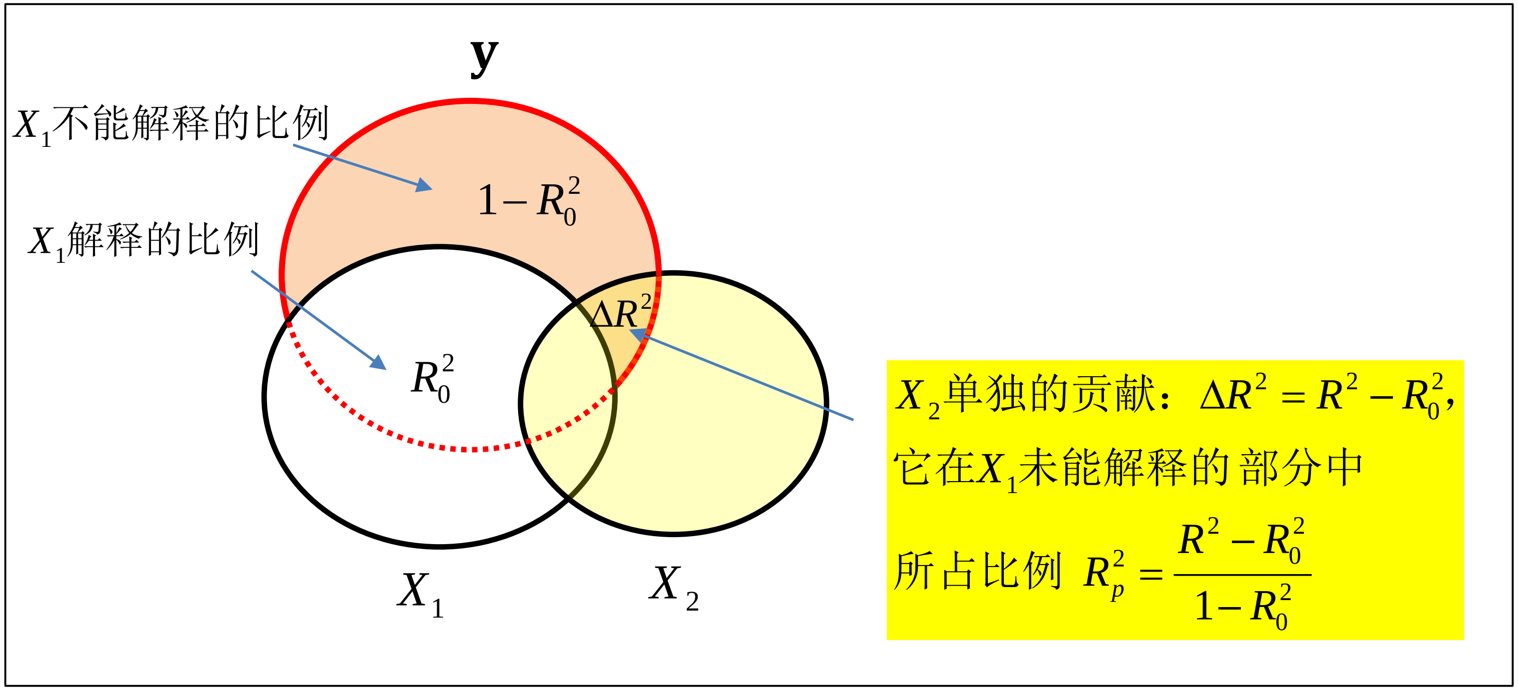
其中$\hat{\boldsymbol{\beta}}_2=(X_2^{\perp\mathrm{T}}X_2^{\perp})^{-1}X_2^{\perp\mathrm{T}}\mathbf{y}$为$\boldsymbol{\beta}_2$的LS估计。

证明：注意到$\hat{\mathbf{y}}=P_{(1,X_1,X_2)}\mathbf{y}=P_{(1,X_1,X_2^{\perp})}\mathbf{y}=\hat{\mathbf{y}}_0+P_{X_2^{\perp}}\mathbf{y}=\hat{\mathbf{y}}_0+X_2^{\perp}\hat{\boldsymbol{\beta}}_2^{\perp}$

$\Rightarrow\hat{\mathbf{y}}-\hat{\mathbf{y}}_0=X_2^{\perp}\hat{\boldsymbol{\beta}}_2^{\perp}$，而$\hat{\sigma}^2=\|\mathbf{y}-\hat{\mathbf{y}}\|^2/(n-p)$，所以

$$F=\frac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2/q}{\|\mathbf{y}-\hat{\mathbf{y}}\|^2/(n-p)}=\frac{\|X_2^{\perp}\hat{\boldsymbol{\beta}}_2\|^2/q}{\hat{\sigma}^2}$$

假设$R^2$是全模型$\mathbf{y}_{n\times 1}=X_{n\times p}\boldsymbol{\beta}_{p\times 1}+\boldsymbol{\varepsilon}$下的决定系数，$R_0^2$是零模型下的决定系数，前面我们对一般线性假设定义了$R_p^2=\dfrac{R^2-R_0^2}{1-R_0^2}$代表$\mathbf{y}$中$X_1$不能解释的部分中，$X_2$所能解释的比例。命题6说明$R_p^2$实际上是部分/偏决定系数（控制$X_1$后,$\mathbf{y}$与$X_2$之间的决定系数）。



命题6. 正态模型$\mathbf{y}_{n\times 1}=X_{n\times p}\boldsymbol{\beta}_{p\times 1}+\boldsymbol{\varepsilon}$中, $X=(\mathbf{1},X_1,X_2),\boldsymbol{\beta}=(\beta_0,\boldsymbol{\beta}_1^{\mathsf{T}},\boldsymbol{\beta}_2^{\mathsf{T}})$,

$H_0:\boldsymbol{\beta}_2=\mathbf{0}_{q\times 1}$的F检验统计量$F=\dfrac{n-p}{q}\times\dfrac{R_p^2}{1-R_p^2}\overset{H_0}{\sim}F_{q,n-p}$,

其中$R_p^2=\dfrac{S_{y2\bullet 1}S_{22\bullet 1}^{-1}S_{2y\bullet 1}}{S_{yy\bullet 1}}\overset{\Delta}{=}R_{y\mathbf{x}_2\bullet\mathbf{x}_1}^2$为控制$X_1$时$\mathbf{y}$与$X_2$的部分/偏决定系数。

证明：$\hat{\mathbf{y}}-\hat{\mathbf{y}}_0=X_2^{\perp}\hat{\boldsymbol{\beta}}_2\Rightarrow$

$$\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2=\mathbf{y}^{\mathsf{T}}X_2^{\perp}(X_2^{\perp\mathsf{T}}X_2^{\perp})^{-1}X_2^{\perp\mathsf{T}}\mathbf{y}=(n-1)S_{y2\bullet 1}S_{22\bullet 1}^{-1}S_{2y\bullet 1},$$

另外, $\|\mathbf{y}-\hat{\mathbf{y}}_0\|^2=\|\mathbf{y}-\mathbf{1}\bar{y}\|^2-\|\hat{\mathbf{y}}_0-\mathbf{1}\bar{y}\|^2$

$$=(n-1)S_{yy}-(n-1)S_{y1}S_{11}^{-1}S_{1y}=(n-1)S_{yy\bullet 1}$$

所以$R_p^2=\dfrac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2}{\|\mathbf{y}-\hat{\mathbf{y}}_0\|^2}=\dfrac{S_{y2\bullet 1}S_{22\bullet 1}^{-1}S_{2y\bullet 1}}{S_{yy\bullet 1}}$

R输出结果summary中包括了两种重要情形：

- 回归方程的显著性F检验：检验所有自变量系数是否显著
- t检验：检验单个回归系数是否显著

# 特例：回归方程的显著性检验和t检验

$$F=\frac{n-p}{p-1}\times\frac{R^2}{1-R^2}\overset{H_0:\mathbf{b}=0}{\sim}F_{p-1,n-p}$$

回归方程的显著性检验：

假设模型 $\mathbf{y}=X\boldsymbol{\beta}+\boldsymbol{\varepsilon}=\mathbf{1}\beta_0+Z\mathbf{b}+\boldsymbol{\varepsilon},\quad \boldsymbol{\varepsilon}\sim N(0,\sigma^2 I_n),\quad H_0:\mathbf{b}=\mathbf{0},$

回归方程的显著性$F$检验：

$$F\overset{命题5}{=}\frac{\|Z_c\hat{\mathbf{b}}\|^2}{(p-1)\hat{\sigma}^2}\overset{命题4}{=}\frac{n-p}{p-1}\times\frac{R^2}{1-R^2}\sim_{H_0}F_{p-1,n-p}$$

$H_0$下模型为$\mathbf{y}=\mathbf{1}\beta_0+\boldsymbol{\varepsilon}$,
决定系数$R_0^2=1$, 0

其中$Z_c$为Z的中心化，$\hat{\mathbf{b}}$，$\hat{\sigma}^2$为LS估计，$R^2$为决定系数。

$F\geq F_{p-1,n-p}(1-\alpha)$时在水平$\alpha$下否定原假设.

$t$ = 估计值 ÷ 标准差 $\sim$ $t_{n-p}$

单个变量的显著性检验:

假设模型$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{1}\beta_0 + \mathbf{x}_1\beta_1 + \ldots + \mathbf{x}_{p-1}\beta_{p-1} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_n)$,

$H_0 : \beta_k = 0, 1 \le k \le p-1$

记$\mathbf{x}_k^{\perp} = \mathbf{x}_k - P_{X_{(-k)}}\mathbf{x}_k$，$X_{(-k)}$为$X$除第$k$列之外的构成的矩阵，记

$r_p = r_{yx_k \bullet \text{其它}}$为$y$与$x_k$偏相关系数。$H_0$的$F$检验:

$$F \overset{\text{命题5}}{=} \frac{\|\mathbf{x}_k^{\perp}\hat{\beta}_k\|^2}{\hat{\sigma}^2} \overset{\text{命题4}}{=} (n-p)\frac{r_p^2}{1-r_p^2} \overset{H_0}{\sim} F_{1,n-p}$$

F检验等价于$t$检验

$$t = \pm\sqrt{F} = \frac{\hat{\beta}_k}{\hat{\sigma} \ / \|\mathbf{x}_k^{\perp}\|} = \sqrt{n-p}\,\frac{r_p}{\sqrt{1-r_p^2}} \sim_{H_0} t_{n-p}\text{。}$$

$\mathrm{var}(\hat{\beta}_k \mid X) = \sigma^2 / \|\mathbf{x}_k^{\perp}\|^2$

例2. 待遇上的性别歧视案例 (数据集salary：alr4).有大学女教师反映女性在待遇上受到了歧视，为此收集了一批数据，包括每个大学教师的工资Salary，性别Sex, 职称Rank和工龄Year。Salary~Sex得到Sex的效应是接近显著的(p=0.07)，但控制其它变量时Sex不再显著(p=0.46):

$$\text{Salary} = \beta_0 + \beta_1 \text{Sex} + \beta_2 \text{Rank} + \beta_3 \text{Year} + \varepsilon$$

```
lm(formula = Salary ~ Sex + Rank + Year, data = salary)

Coefficients:
              Estimate   Std. Error  t value   Pr(>|t|)
(Intercept)   11011.76   966.95      11.388    3.03e-15 ***
Sex           603.77     811.20      0.744     0.46
Rank          4747.18    452.58      10.489    5.18e-14 ***
Year          393.86     74.53       5.285     3.04e-06 ***
---
Residual standard error: 2398 on 48 degrees of freedom
Multiple R-squared: 0.8454,    Adjusted R-squared: 0.8358
F-statistic: 87.51 on 3 and 48 DF,  p-value: < 2.2e-16
```

$t$ = 估计值 ÷ 标准差

**t=603.77/811.20**
**=0.744**

回归方程的显著性F检验

单个回归系数的t检验

在R软件中，summary没有输出的检验需要使用anova函数，比较完全模型和零模型

例如，同时检验Rank, Year的显著性：$H_0:\beta_2=\beta_3=0$。

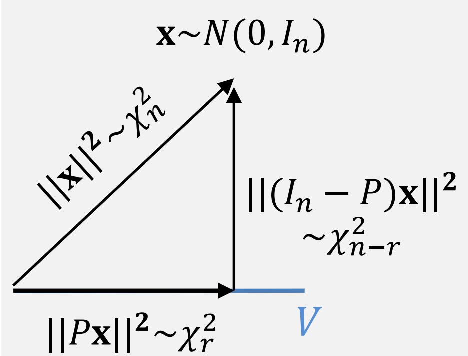
```
> m1  = lm(Salary~Sex+Rank+Year, data=salary)   #完全模型
> m0  = lm(Salary~Sex, data=salary)              #零模型
> anova(m0, m1)  #方差分析anova：比较m0,m1
Analysis of Variance Table

Model 1: Salary ~ Sex
Model 2: Salary ~ Sex + Rank + Year
  Res.Df      RSS        Df    Sum of Sq     F        Pr(>F)
1    50       1671623638
2    48       276016717   2     1395606921  121.35   < 2.2e-16 ***
```

$$F=\frac{n-p}{k}\times\frac{RSS_0-RSS}{RSS}=\frac{52-4}{2}\times\frac{1671623638-276016717}{276016717}=121.35$$

# 总结：$F$检验的构造及零分布

$H_0: \theta = \theta_0$的Wald 检验$W = (\hat{\theta} - \theta_0)^{\mathrm{T}}[\widehat{var}\ (\hat{\theta})]^{-1}(\hat{\theta} - \theta_0)$

$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_n)$，线性假设$H_0: A\boldsymbol{\beta} = \mathbf{0}_{q\times 1}$，$F$检验构造如下：




- Wald 检验：$W = (\mathrm{A}\widehat{\boldsymbol{\beta}})^T[\widehat{var}(\mathrm{A}\widehat{\boldsymbol{\beta}})]^{-1}(\mathrm{A}\widehat{\boldsymbol{\beta}})$
- $F$检验： $F = W/q$
- 由约束最小二乘结果，以投影表示$F$

$$F = \frac{n-p}{q} \times \frac{||\hat{\mathbf{y}} - \hat{\mathbf{y}}_0||^2}{||\mathbf{y} - \hat{\mathbf{y}}||^2}, \quad \hat{\mathbf{y}} = P_V\mathbf{y},\ \hat{\mathbf{y}}_0 = P_{V_0}\mathbf{y}$$

其中 $V = C(X)$和$V_0$ 分别为全模型下和原假设的自变量空间。

- $H_0$成立时，$F \sim F_{q,n-p}$。

- $F$的其它表示：$F = \frac{n-p}{q} \times \frac{RSS_0 - RSS}{RSS} = \frac{n-p}{q} \times \frac{R^2 - R_0^2}{1-R^2} = \frac{n-p}{q} \times \frac{R_p^2}{1-R_p^2}$，

即F检验比较全模型和原假设下的残差平方和或比较决定系数，

$R_p^2 = \frac{R^2 - R_0^2}{1 - R_0^2}$ 为部分决定系数。

# $F$检验所有表示形式，都是对比全模型和零模型下的拟合的差异

$H_0:\boldsymbol{\beta}_2=\mathbf{0}_{q\times 1}$

$F=\frac{n-p}{q}\times\frac{\|\hat{\mathbf{y}}-\hat{\mathbf{y}}_0\|^2}{\|\mathbf{y}-\hat{\mathbf{y}}\|^2}$ 几何解释

$F=\frac{n-p}{q}\times\frac{RSS_0-RSS}{RSS}$ 几何解释,软件使用

$F=\frac{\|X_2^{\perp}\hat{\boldsymbol{\beta}}_2\|^2}{q\hat{\sigma}^2}$ 直观

统一形式

$F=\frac{n-p}{q}\times\frac{R_p^2}{1-R_p^2}$

- $q=1$情形：

$p=2,q=1$: 两个变量的相关性, $F=t^2=(n-2)\frac{r^2}{1-r^2}\overset{H_0}{\sim}F_{1,n-2}$

$p>2,q=1$: 两个变量的偏相关性, $F=t^2=(n-p)\frac{r_p^2}{1-r_p^2}\overset{H_0}{\sim}F_{1,n-p}$

- $q>1$情形：

$q=p-1$: 回归方程的显著性, $F=\frac{n-p}{q}\times\frac{R^2}{1-R^2}\overset{H_0}{\sim}F_{p-1,n-p}$

$1<q<p-1$: $F=\frac{n-p}{q}\times\frac{R_p^2}{1-R_p^2}$

# 单因素方差分析 (one-way anova)

ANOVA：Analysis of variance

方差分析检验多个正态总体均值相等，是试验设计的重要分析方法。
方差分析是线性回归模型中所有自变量为因子变量的特殊情形。

随机化控制试验中，研究对象被随机分配接受$K$种处理(*treatment*)中的一种（比如$K$种药物剂量），测量响应，假设各个组的响应服从方差相同但均值可能不同的正态分布，检验各组均值相等，称为单因素方差分析。

$$\bar{y}_{k\bullet} = \sum_{i=1}^{n_k} y_{ki} / n_k,$$

$$\bar{y}_{\bullet\bullet} = \left( \sum_{k=1}^{K} \sum_{i=1}^{n_k} y_{ki} \right) / n,$$

| 处理 | $N(\mu_1, \sigma^2)$ | ... | $N(\mu_K, \sigma^2)$ |
|---|---|---|---|
| 响应 | $y_{11}, y_{12}, \ldots, y_{1n_1}$ | ... | $y_{K1}, y_{K2}, \ldots, y_{Kn_K}$ |
| 平均 | $\bar{y}_{1\bullet}$ | ... | $\bar{y}_{K\bullet}$ |

ANOVA 模型

单因素方差法分析模型（双下标形式）：

$$y_{ki} = \mu_k + \varepsilon_{ki}, \quad \varepsilon_{ki} \ iid \sim N(0, \sigma^2), \ i = 1, \ldots, n_k; k = 1, .., K.$$

$H0: \mu_1 = \ldots = \mu_K$,写成通常的线性模型形式：

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{g}_1 \mu_1 + \ldots + \mathbf{g}_K \mu_K + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I_n),$$

其中$\mathbf{g}_1, \ldots, \mathbf{g}_K$为各组的示性表示，$X = (\mathbf{g}_1, \ldots, \mathbf{g}_K)$. 具体如下：

$$
\begin{array}{c}
\text{第一组} \\ \\ \text{第二组} \\ \\ \text{第}K\text{组}
\end{array}
\begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{K1} \\ \vdots \\ y_{Kn_K} \end{pmatrix}
=
\begin{pmatrix}
1 & 0 & \dots & 0 \\
\vdots & \vdots & \dots & 0 \\
1 & 0 & \dots & 0 \\
0 & 1 & \dots & 0 \\
\vdots & \vdots & & \\
0 & 1 & \dots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & 1 \\
\vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_K \end{pmatrix}
+
\begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \\ \vdots \\ \varepsilon_{K1} \\ \vdots \\ \varepsilon_{Kn_K} \end{pmatrix}
=
\underbrace{\begin{pmatrix}
\mathbf{1}_{n_1} & \mathbf{0} & \dots & \mathbf{0} \\
\mathbf{0} & \mathbf{1}_{n_2} & \dots & \mathbf{0} \\
\vdots & \vdots & \dots & \vdots \\
\mathbf{0} & \mathbf{0} & \dots & \mathbf{1}_{n_K}
\end{pmatrix}}_{(\mathbf{g}_1,\ \mathbf{g}_2, \dots, \mathbf{g}_K)}
\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_K \end{pmatrix}
+ \boldsymbol{\varepsilon},
$$

下面求$H_0 : \mu_1 = \dots = \mu_K$的$F$检验： $F = \dfrac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2 / q}{\|\mathbf{y} - \hat{\mathbf{y}}\|^2 / (n-p)}$， $p = K, q = K-1$

$$\hat{\mathbf{y}} = P_{\mathbf{g}_1}\mathbf{y} + \dots + P_{\mathbf{g}_K}\mathbf{y} = \mathbf{g}_1\bar{y}_{1\bullet} + \dots + \mathbf{g}_K\bar{y}_{K\bullet} = (\bar{y}_{1\bullet}, \dots \bar{y}_{1\bullet}, \bar{y}_{2\bullet}, \dots, \bar{y}_{2\bullet}, \dots\dots, \bar{y}_{K\bullet}, \dots \bar{y}_{K\bullet})^{\mathsf{T}}$$

$H_0$下模型为： $\mathbf{y} = (\mathbf{g}_1 + \dots + \mathbf{g}_K)\mu_1 + \boldsymbol{\varepsilon} = \mathbf{1}\mu_1 + \boldsymbol{\varepsilon}$, $\quad V_0 = L(\mathbf{1})$.

$$\hat{\mathbf{y}}_0 = P_{V_0}\mathbf{y} = (\bar{y}_{\bullet\bullet}, \dots\dots, \bar{y}_{\bullet\bullet})^{\mathsf{T}} = \mathbf{1}\bar{y}_{\bullet\bullet}$$

容易验证$F = \dfrac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2 / q}{\|\mathbf{y} - \hat{\mathbf{y}}\|^2 / (n-p)} = \dfrac{SS_B / (K-1)}{SS_W / (n-K)}$（下页）

命题7. 单因素方差分析,零假设$H_0: \mu_1 = \ldots = \mu_K$的F检验

$$F = \frac{SS_B/(K-1)}{SS_W/(n-K)} \sim_{H_0} F_{K-1,n-K}.,$$

- $SS_B = \|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2 = \sum_{k=1}^{K}\sum_{i=1}^{n_k}(\bar{y}_{k\bullet} - \bar{y}_{\bullet\bullet})^2$

  组间平方和(回归平方和),度量了各组的中心 $\bar{y}_{k\bullet}, k = 1,\ldots,K$之间的差异。

- $SS_W = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 = \sum_{k=1}^{K}\sum_{i=1}^{n_k}(y_{ki} - \bar{y}_{k\bullet})^2,$

  组内平方和(残差平方和),度量了组内样本的差异。

- $SS_{总} = SS_B + SS_W$, B : Between - group, W : Within - group

注：$F$的主要部分$SS_B$度量了$K$个中心的差异($SS_W$仅用来标准化）。

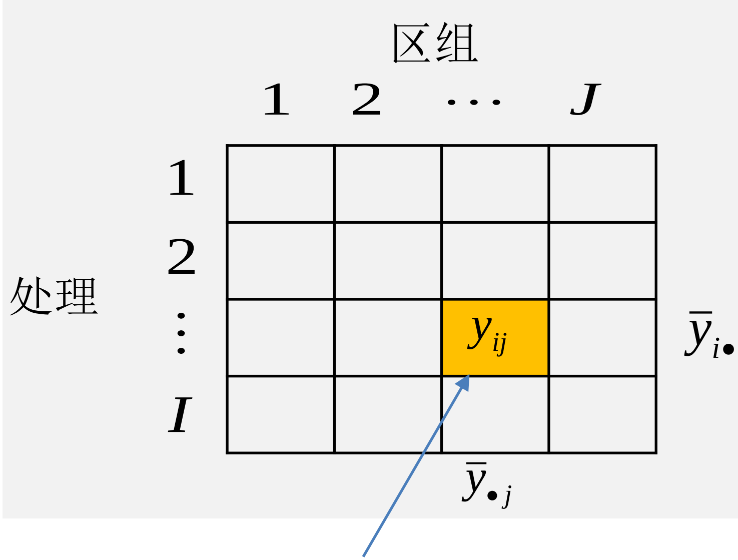
特别地，当$K = 2$（两正态），$F = t^2$, $t \sim_{H_0} t_{n-2}$.

R命令： > aov(响应 ~ 组别, data =… )
> lm(响应 ~ 组别, data =… )

# 两因素方差分析

**随机区组设计(Randomized block design)**

随机化试验设计中，当研究对象差异较大时，为了避免处理组与对照组出现严重的不均衡，通常先将研究对象分成内部个体相似的区组或层(blocks/strata)，在区组内随机化分配处理。关心的问题是各个处理是否存在差异。决定分组的因素/因子有两个：处理和区组。假设两个因子各有$I,J$个水平，交叉分类得到$I\times J$表，只考虑格子内只有一次测量的情形




$\alpha_i$: 处理$i$的效应

$\beta_j$: 区组$j$的效应

$H_0$: $\alpha_1 = \ldots = \alpha_I$

模型：$y_{ij} \sim N(\mu_{ij}, \sigma^2)$，假设可加模型：$\mu_{ij} = \mu + \alpha_i + \beta_j$, 或

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}，\ \varepsilon_{ij} \sim N(0, \sigma^2)$$

## 两因素方差分析F检验

命题8. $I$个处理水平, $J$个区组, 假设$(i,j)$水平下的测量$y_{ij}$满足可加模型

$$y_{ij}=\mu+\alpha_i+\beta_j+\varepsilon_{ij},\ \varepsilon_{ij}\ \text{iid}\sim N(0,\sigma^2),$$ （$\alpha$'s和$\beta$'s各有一个约束）

$H_0:\alpha_1=\ldots=\alpha_I=0$ 的$F$检验为:

$$F=\frac{SS_{处理}/(I-1)}{SS_{组内}/(I-1)(J-1)}\overset{H_0}{\sim}F_{I-1,(I-1)(J-1)}$$

- $SS_{处理}=\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2$ ,K个处理之间的差异
- $SS_{组内}=\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij}-\bar{y}_{i\bullet}-\bar{y}_{\bullet j}+\bar{y}_{\bullet\bullet})^2$, 残差平方和

行平均$\bar{y}_{i\bullet}=\sum_{j=1}^{J}y_{ij}/J$,

列平均 $\bar{y}_{\bullet j}=\sum_{i=1}^{I}y_{ij}/I$,

总平均$\bar{y}_{\bullet\bullet}=\sum_{j=1}^{J}\sum_{j=1}^{J}y_{ij}/IJ$

Key: $\mu_{ij}=\mu+\alpha_i+\beta_j$ 表明$\mu_{ij}-\mu_{i'j}=\alpha_i-\alpha_{i'}$不依赖于区组$j$, 即两个处理$i$, $i'$的差异在各个区组内相同, 因此处理平方和与区组效应$\beta's$无关:

$$SS_{处理}=\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2=J\sum_{i=1}^{I}(\alpha_i-\bar{\alpha}+\bar{\varepsilon}_{i\bullet}-\bar{\varepsilon}_{\bullet\bullet})^2$$

$\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet}$代表了处理$i$的效应，要点在于不同区组的响应可以累加$\bar{y}_{i\bullet}$,
但其中包含的累加的区组效应与$\bar{y}_{\bullet\bullet}$中包含的区组效应抵消了。
列联表分析中类似的思想可得Cochran - Mantel - Haenszel检验。

# 配对$t$检验

**配对设计**

配对设计(paired design)：两个相像的研究对象（比如双胞胎），随机选取其中一人服用药物（处理），另一个服用安慰剂（对照）。

| pair | 1 | 2 | … | $J$ | 平均 |
|---|---|---|---|---|---|
| 对照 | $y_{11}$ | $y_{12}$ | … | $y_{1J}$ | $\bar{y}_{1\bullet}$ |
| 处理 | $y_{21}$ | $y_{22}$ | … | $y_{2J}$ | $\bar{y}_{2\bullet}$ |
| 差 | $z_1$ | $z_2$ | … | $z_J$ | $\bar{z}$ |

第$j$对样本组成一个区组，响应：$(y_{1j}, y_{2j})$，下标2表示处理,1表示对照。

假设两因素方差分析模型

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma^2), \underline{i = 1,2;}\ j = 1, \ldots, J$$

其中$\alpha_1 = 0$，$\alpha_2$代表药效。

配对设计是最简单的区组设计，每一个pair就是一个区组。区组内的两个研究对象相似， 但接受不同的处理。可以验证命题8当$I = 2$ 时，$F = (t_{pair})^2$，$t_{pair}$具有如下形式

**成对t检验 (paired t-test/dependent t-test)**

命题9.配对样本$(y_{1i}, y_{2i})$，$i=1,\dots,n$ iid，令$z_i = y_{2i} - y_{1i}$，　$\alpha_2 = E(z_i)$，$H_0: \alpha_2 = 0$的成对$t$检验

$$t_{\text{pair}} = \frac{\bar{z}}{s_z/\sqrt{n}} = \frac{(\bar{y}_{2\bullet} - \bar{y}_{1\bullet})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_{2i} - y_{1i} - (\bar{y}_{2\bullet} - \bar{y}_{1\bullet}))^2/n}} \overset{H_0}{\sim} t_{n-1}$$

同一个pair内的两个个体相似(匹配、相依)，将它们作为独立样本进行两样本t检验是错误的做法.

# 置信域和同时置信区间

Tukey's quote: *Far better an approximate answer to the right question, which is often vague* (同时置信区间), *than an exact answer to the wrong question, which can always be made precise*（置信域）.

回归系数的置信域

正态线性模型 $\mathbf{y}_{n\times1} = X_{n\times p}\boldsymbol{\beta}_{p\times1} + \boldsymbol{\varepsilon}_{n\times1}$， $\boldsymbol{\varepsilon} \sim N(0,\sigma^2 I_n)$

由定理2，对任何行满秩$q\times p$常数矩阵$A=(\mathbf{a}_1,\ldots,\mathbf{a}_q)^\mathsf{T}$，

$$\frac{(A\hat{\boldsymbol{\beta}}-A\boldsymbol{\beta})^\mathsf{T}(A(X^\mathsf{T}X)^{-1}A^\mathsf{T})^{-1}(A\hat{\boldsymbol{\beta}}-A\boldsymbol{\beta})}{q\hat{\sigma}^2} \sim F_{k,n-p},$$

称为枢轴量（分布与参数无关），基于此我们构建置信域：

$\boldsymbol{\theta} = A\boldsymbol{\beta}$的(即$q$个参数$\theta_1=\mathbf{a}_1^\mathsf{T}\boldsymbol{\beta},\ldots,\theta_q=\mathbf{a}_q^\mathsf{T}\boldsymbol{\beta}$的)置信度为$1-\alpha$的置信椭球

$$C=\left\{\boldsymbol{\theta}\in R^q:\ \frac{(\boldsymbol{\theta}-A\hat{\boldsymbol{\beta}})^\mathsf{T}\left[A(X^\mathsf{T}X)^{-1}A^\mathsf{T}\right]^{-1}(\boldsymbol{\theta}-A\hat{\boldsymbol{\beta}})}{q\hat{\sigma}^2}\le F_{q,n-p}(1-\alpha)\right\},$$

注意$C$是$R^q$中的椭球区域，$P(C)=1-\alpha$。

注：当参数个数为$q=1$的时候，上述置信椭球为通常的置信区间：

- 若$A=\mathbf{a}^{\top}$（行向量），$\mathbf{a}^{\top}\boldsymbol{\beta}$的$(1-\alpha)100\%$置信区间:

$$\left\{\mathbf{a}^{\top}\boldsymbol{\beta}:|\mathbf{a}^{\top}\boldsymbol{\beta}-\mathbf{a}^{\top}\hat{\boldsymbol{\beta}}|\le t_{n-p}(\alpha/2)\hat{\sigma}\sqrt{\mathbf{a}^{\top}(X^{\top}X)^{-1}\mathbf{a}}\right\}$$

- 若$A=(0,...,1,...0)^{\top}$时，$\beta_k$的置信区间：

$$\left\{\beta_k:|\beta_k-\hat{\beta}_k|\le t_{n-p}(\alpha/2)\hat{\sigma}/\|\mathbf{x}_k^{\perp}\|\right\}$$

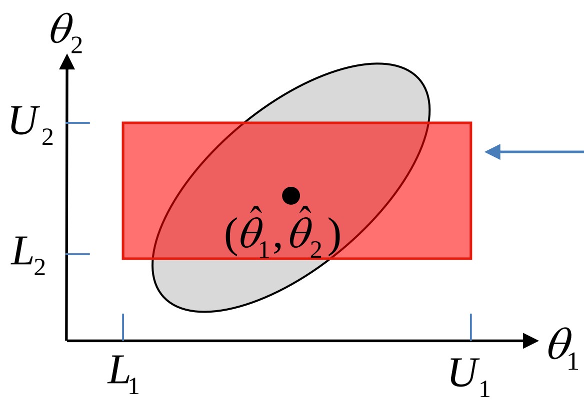
参数$\theta_1=\mathbf{a}_1^{\top}\boldsymbol{\beta},...,\theta_q=\mathbf{a}_q^{\top}\boldsymbol{\beta}$的置信椭球不容易理解和解释，能否构建$q$个参数的方型置信域,称为同时置信区间(simultaneous confidence interval)?

$$\left\{(\theta_1,...\theta_q)^{\top}: L_j\le\theta_j\le U_j, j=1,...,q,(L_j+U_j)/2=\hat{\theta}_j=\mathbf{a}_j{}^{\top}\hat{\boldsymbol{\beta}}\right\}$$

由$q$个区间构成:$\otimes\{\theta_j: L_j\le\theta_j\le U_j\}$，容易解释和应用。。

可以想象，每个边际区间都要比只考虑一个参数的时候稍长，
我们希望同时置信区间覆盖参数$\theta_1,\ldots\theta_q$的概率等于给定的
置信水平$1-\alpha$,但实际上即使在联合正态假设下这也难以做到，
我们只能粗略地控制覆盖概率不低于置信水平$1-\alpha$(即偏保守).

置信椭球 “wrong” question, accurate answer




同时置信区间:

$L_1 \le \theta_1 \le U_1$
$L_2 \le \theta_2 \le U_2$

right question, approximate answer

常见的方法有：Bonferroni同时置信区间,Sheffe同时置信区间,
方差分析模型的Tukey同时置信区间。

Bonferroni 不等式

Bonferroni 不等式：$P\left(\bigcup_{j=1}^{q} \mathrm{E}_j\right) \le \sum_{j=1}^{q} P(\mathrm{E}_j)$

假设参数$\theta_j$的置信区间为$E_j, \theta_1, \ldots, \theta_q$的同时置信区间为$\bigcap E_j$，我们希望

$$P\left(\bigcap_{j=1}^{q} E_j\right) \ge 1-\alpha \quad \text{(给定的置信水平）}$$

假设每个区间覆盖真参数的概率为$1-\gamma$,即$P(E_j)=1-\gamma$,由Bonferroni不等式

$$P\left(\bigcap_{j=1}^{q} E_j\right) = 1-P(\bigcup_{j=1}^{q} \bar{E}_j) \ge 1-\sum P(\bar{E}_j) = 1-q\gamma$$

取$\gamma=\alpha/q$,即每个区间$E_j$的置信度取$1-\alpha/q$即可保证$\bigcap E_j$的覆盖率不小于$1-\alpha$.

若$E_j$独立，则$P\left(\bigcap E_j\right)=\prod P(E_j)=(1-\alpha/q)^q \approx 1-q\alpha$,此时Bonferroni不等式几乎是等式，即Bonferroni不等式/置信区间几乎是不可改进的。

Bonferroni同时置信区间

$A\boldsymbol{\beta}$的Bonferroni同时置信区间：

$$\bigcap_{j=1}^{q} E_j = \bigcap_{j=1}^{q} \left\{ \mathbf{a}_j^{\mathsf{T}}\boldsymbol{\beta}: |\mathbf{a}_j^{\mathsf{T}}\boldsymbol{\beta} - \mathbf{a}_j^{\mathsf{T}}\hat{\boldsymbol{\beta}}| \le t_{n-p}\left(\boxed{\frac{\alpha}{2q}}\right)\hat{\sigma}\sqrt{\mathbf{a}_j^{\mathsf{T}}(X^{\mathsf{T}}X)^{-1}\mathbf{a}_j} \right\}$$

称为是$\mathbf{a}_1^{\mathsf{T}}\boldsymbol{\beta}, \ldots, \mathbf{a}_q^{\mathsf{T}}\boldsymbol{\beta}$的置信度为$1-\alpha$的Bonferroni同时置信区间

每个置信区间的置信度为$1-\frac{\alpha}{q} > 1-\alpha$，

比置信度$1-\alpha$ 的区间稍宽

多重检验

多重假设检验(multiple testing)同时检验$q$个零假设（对应于同时置信区间），每次检验的错误率假设为$\beta$，多次检验发生错误的概率会增加（inflation of typeI error rate)。我们希望所有零假设成立的时候，否定至少一个零假设的概率(即$I$型错误率）不大于给定的水平$\alpha$。
最常用的方法是Bonferroni校正,即每个假设检验的水平选取为$\beta = \alpha / q$.

应用/动机：事后分析

事后分析(post - hoc analysis)：

- 单因素方差分析检验$\mu_1 = \ldots = \mu_g$，F检验称为omnibus检验(一揽子检验)
- 如果结果显著，我们希望发现到底是哪些组之间不同，即事后分析
- 事后分析需要给出$g(g-1)/2$个差别$\mu_i - \mu_j, 1 \le i < j \le g$的同时置信区间，或等价地检验$g(g-1)/2$个原假设$H_0^{(ij)}$：$\mu_i = \mu_j$, $1 \le i < j \le g$.
- 显然各个检验/区间不独立，Bonferroni方法过于保守，Tukey提出了一种 基于学生化极差分布的同时置信区间。

应用:全基因组关联分析

全基因组数十万、百万个基因，哪些与疾病关联？
不可能同时检验，通常逐个检验每个基因与疾病的关联性，这是多重检验，需要控制一型错误率。